Attached you will find a 100011-word excerpt from the novel “A Tale of Two Cities” by Charles Dickens (A\_tale\_of\_two\_cities.100kwords.original.txt).

You will also find an obfuscated version where each instance of the words “a” and “the” have been replaced by the words “a” or “the” (A\_tale\_of\_two\_cities.100kwords.obfuscated.txt).

Your task is to resolve the “a/the” ambiguity in the latter version as best as possible via a computer program. Note this program should not access any part of the novel “A Tale of Two Cities” to accomplish this; i.e. it cannot “train on” or “memorize” this particular novel.  You are otherwise free to use whatever approach you see fit.

In particular, please provide

1. A working program which takes “A\_tale\_of\_two\_cities.100kwords.obfuscated.txt “ as input and generates an output file where each instance of the word “a” and “the”  is disambiguated.
2. The output of this program in two formats
   * Plain text format
   * Three column TSV with one word per line where column-1 contains original word, column-2 contains obfuscated word, and column-3 contains disambiguated word. Words other than a/the also needs to be present.
3. An assessment of the output of the program (i.e. the fraction of “a/the” instances correctly resolved).
4. A description of how you approached the problem, how your program works and how much time you spent.